



Implementasi Metode XGBoost dan SHAP untuk Klasifikasi dan Analisis Faktor Risiko Penyakit Diabetes Mellitus

Fatika Lovina Febrianti¹, Indah Manfaati Nur², M. Al Haris³, Saeful Amri⁴

^{1,2,3,4}Statistika, Universitas Muhammadiyah Semarang

¹fatikalovina@gmail.com

²indahmnur@unimus.ac.id

³alharis@unimus.ac.id

⁴saefulamri@unimus.ac.id

Corresponding author email: indahmnur@unimus.ac.id

Abstract: *Diabetes Mellitus (DM) is a non-communicable disease with a steadily increasing prevalence. Early detection and a clear understanding of risk factors are crucial to prevent more severe complications. This study aims to develop a diabetes risk classification model and identify the most influential factors affecting the prediction results. The methods used include the Extreme Gradient Boosting (XGBoost) algorithm for classification and Shapley Additive Explanations (SHAP) to evaluate the contribution of each variable. The dataset employed is the Early Stage Diabetes Risk Prediction Dataset from the UCI Machine Learning Repository, consisting of 520 observations and 16 variables. The results show that the XGBoost model performs well in classifying diabetes risk, achieving an accuracy of 98.7%. SHAP analysis identifies polyuria, polydipsia, and gender as the most influential variables in determining diabetes risk classification. The integration of XGBoost and SHAP not only provides highly accurate predictions but also offers transparent explanations of the model's decisions, making it a promising tool for early detection and data-driven medical decision support.*

Keywords: *Classification, Diabetes Mellitus, Risk Factors, SHAP, XGBoost*

Abstrak: Diabetes Mellitus (DM) adalah penyakit tidak menular dengan prevalensi yang terus meningkat. Deteksi dini dan pemahaman terhadap faktor risiko menjadi langkah penting untuk mencegah komplikasi yang lebih serius. Penelitian ini bertujuan membangun model klasifikasi risiko diabetes serta mengidentifikasi faktor-faktor yang paling berpengaruh terhadap hasil prediksi. Metode yang digunakan adalah algoritma *Extreme Gradient Boosting* (XGBoost) untuk klasifikasi dan *Shapley Additive Explanations* (SHAP) untuk menilai pengaruh masing-masing variabel. Dataset yang digunakan adalah *Early Stage Diabetes Risk Prediction Dataset* dari *UCI Machine Learning Repository* yang berisi 520 observasi dengan 16 variabel. Hasil penelitian menunjukkan bahwa model XGBoost mampu mengklasifikasikan risiko diabetes dengan baik, dengan akurasi 98,7%. SHAP mengidentifikasi gejala *polyuria*, *polydipsia*, serta faktor *gender* sebagai variabel yang paling berkontribusi terhadap klasifikasi risiko diabetes. Kombinasi XGBoost dan SHAP tidak hanya mampu memberikan prediksi yang akurat, tetapi juga menjelaskan alasan di balik keputusan model, sehingga dapat menjadi alat bantu yang potensial dalam deteksi dini dan pengambilan keputusan medis berbasis data.

Kata kunci: Diabetes Mellitus, Faktor Risiko, Klasifikasi, SHAP, XGboost

I. PENDAHULUAN

Diabetes Mellitus (DM) merupakan salah satu penyakit tidak menular yang terus menunjukkan peningkatan prevalensi dari tahun ke tahun, baik secara global maupun di Indonesia. Berdasarkan data *IDF Diabetes Atlas*, tercatat sekitar 537 juta orang dewasa di dunia yang mengidap diabetes, dan angka ini diproyeksikan meningkat menjadi 783 juta pada tahun 2045. Di Indonesia sendiri, prevalensi DM pada kelompok usia 20–79 tahun mencapai 10,6% atau sekitar 19,5 juta orang. Kondisi ini menempatkan Indonesia sebagai salah satu dari sepuluh negara dengan jumlah penderita diabetes tertinggi di dunia. Selain itu, lebih dari separuh kasus tersebut belum terdiagnosis, sehingga risiko komplikasi serius kerap tidak terdeteksi sejak dini [1]. Tingginya jumlah kasus yang tidak terdeteksi menunjukkan perlunya pendekatan deteksi dini yang lebih efektif.

Gejala awal diabetes sering kali tidak dikenali oleh penderita. Misalnya, gejala seperti sering buang air kecil (*polyuria*), rasa haus berlebihan (*polydipsia*), dan penurunan berat badan mendadak dapat dianggap sepele atau tidak disadari oleh penderita. Keterlambatan dalam mengenali dan menangani gejala-gejala tersebut dapat menyebabkan risiko komplikasi seperti penyakit jantung, gagal ginjal, dan gangguan saraf [2]. Identifikasi gejala dan faktor risiko sejak awal menjadi langkah penting dalam mencegah perkembangan penyakit.

Kemajuan teknologi digital dan ketersediaan data medis telah membuka peluang baru dalam pengembangan sistem deteksi dini berbasis *machine learning*. Berbagai algoritma telah digunakan untuk memprediksi risiko diabetes, seperti *Logistic Regression*, *Naïve Bayes*, *Random Forest*, dan *Support Vector Machine* (SVM) [3][4]. Hasil penelitian tersebut menunjukkan bahwa pendekatan berbasis algoritma mampu menghasilkan akurasi prediksi yang cukup baik. Salah satu algoritma yang menunjukkan performa tinggi dalam tugas klasifikasi adalah *Extreme Gradient Boosting* (XGBoost). Beberapa penelitian telah membuktikan efektivitas XGBoost dalam berbagai analisis kesehatan, seperti pada klasifikasi penyakit Parkinson dengan akurasi sebesar 85,60% [5], klasifikasi obesitas pada populasi dewasa dengan akurasi 92% [6], serta klasifikasi status stunting pada anak balita dengan akurasi 86% [7]. Meskipun demikian, tantangan utama yang sering ditemui adalah aspek interpretabilitas model yang masih rendah.

Namun demikian, tantangan utama yang masih sering ditemukan adalah kurangnya kemampuan interpretasi dari model yang digunakan. Beberapa algoritma, seperti XGBoost, dikenal memiliki performa klasifikasi yang sangat tinggi, tetapi bersifat *black-box* model, sehingga sulit dipahami bagaimana setiap fitur memengaruhi hasil prediksi. Padahal dalam konteks kesehatan, interpretabilitas menjadi hal yang sangat krusial. Tenaga medis membutuhkan informasi yang dapat dijelaskan secara logis dan transparan agar dapat mendukung pengambilan keputusan klinis secara akurat dan bertanggung jawab.

Metode *Shapley Additive Explanations* (SHAP) dikembangkan untuk mengatasi keterbatasan interpretasi pada model *machine learning*. SHAP memberikan informasi kuantitatif mengenai kontribusi setiap fitur terhadap hasil prediksi serta menyediakan visualisasi yang memudahkan pemahaman. Penggunaan SHAP memungkinkan hasil klasifikasi tidak hanya akurat, tetapi juga dapat dijelaskan secara rinci, sehingga meningkatkan kepercayaan pengguna terhadap model. Beberapa penelitian sebelumnya telah menerapkan metode SHAP dalam berbagai bidang kesehatan, seperti penyakit ginjal kronis dan kanker paru-paru [8][9]. Namun, penerapannya pada klasifikasi risiko diabetes, khususnya dalam konteks pemahaman faktor risiko secara menyeluruh, masih belum banyak dilakukan.

Penelitian ini bertujuan untuk membangun model klasifikasi risiko diabetes menggunakan algoritma XGBoost serta menganalisis kontribusi masing-masing variabel terhadap prediksi menggunakan metode SHAP. Model yang dihasilkan diharapkan dapat memberikan prediksi yang akurat sekaligus transparan, sehingga bermanfaat sebagai alat bantu dalam sistem deteksi dini dan pengambilan keputusan berbasis data di bidang kesehatan.

II. METODE PENELITIAN

Penelitian ini menggunakan data sekunder yaitu *Early Stage Diabetes Risk Prediction Dataset*. Dataset diperoleh dari *UCI Machine Learning Repository* dan berisi 520 observasi dengan 16 variabel prediktor serta satu variabel target berupa status risiko diabetes (positif atau negatif). Variabel prediktor mencakup data demografis, yaitu usia dan jenis kelamin, serta gejala klinis yang umum dialami oleh penderita diabetes, yaitu *polyuria*, *polydipsia*, *sudden weight loss*, *weakness*, *polyphagia*, *genital thrush*, *visual blurring*, *itching*, *irritability*, *delayed healing*, *partial paresis*, *muscle stiffness*, *alopecia*, dan *obesity*.

2.1. *Extreme Gradient Boosting* (XGBoost)

XGBoost adalah pengembangan dari metode *gradient boosting* yang diperkenalkan sebagai algoritma yang mampu melakukan pemodelan prediktif secara lebih efisien dan akurat dibandingkan metode boosting konvensional [10]. Algoritma ini merupakan penyempurnaan dari *Gradient Boosting Decision Tree* (GBDT), dengan kemampuan pemrosesan paralel, regulasi untuk menghindari *overfitting*, dan teknik pruning berbasis nilai *loss* yang diperoleh dari setiap *node* [11]. Algoritma ini bekerja dengan cara menambahkan pohon keputusan secara bertahap untuk meminimalkan fungsi objektif, yang terdiri dari fungsi kerugian dan fungsi regulasi. Fungsi objektif dalam XGBoost dapat dituliskan sebagai berikut [12]:

$$L(\emptyset) = \sum_i l(y_i, \hat{y}_i) + \sum_k \Omega(f_k) \quad (1)$$

Dengan $l(y_i, \hat{y}_i)$ adalah fungsi kerugian (*loss function*) yang mengukur selisih antara \hat{y}_i (nilai prediksi) dan y_i (nilai aktual). Sedangkan $\Omega(f_k)$ merupakan fungsi regulasi yang mengontrol kompleksitas model guna mencegah *overfitting* pada model dan dinyatakan sebagai [13]:

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \|w\|^2 \quad (2)$$

Mengacu pada (2), γ berperan sebagai koefisien penalti yang mengatur kompleksitas model, dan λ adalah parameter penalti yang diterapkan pada bobot daun ω .

2.2. Shapley Additive Explanations (SHAP)

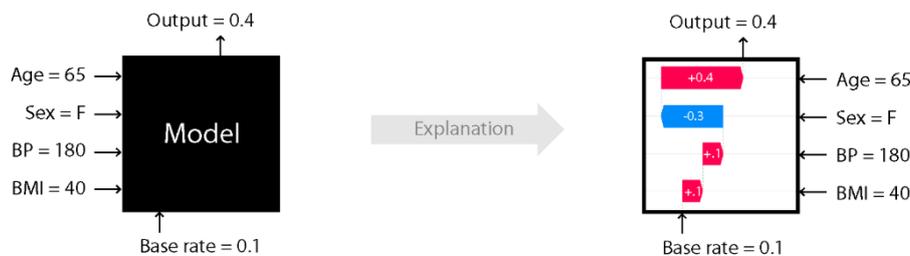
SHAP (*Shapley Additive Explanations*) adalah metode interpretasi model yang berguna untuk memahami kontribusi masing-masing fitur terhadap prediksi. SHAP didasarkan pada teori permainan kooperatif, yaitu *Shapley*, di mana setiap fitur dianggap sebagai “pemain” yang turut memberikan kontribusi pada keputusan model. Model SHAP dijelaskan dalam bentuk fungsi aditif berikut [14]:

$$g(\mathbf{z}) = \phi_0 + \sum_{j=1}^M \phi_j z_j \quad (3)$$

Di mana $g(\mathbf{z})$ adalah model interpretasi, $\mathbf{z} \in \{0,1\}^M$ dengan nilai 1 menunjukkan bahwa suatu variabel diperhitungkan dan nilai 0 jika tidak. M adalah jumlah total fitur dalam model, ϕ_0 merupakan nilai dasar (*baseline value*) dari model klasifikasi, dan $\phi_j \in R$ adalah nilai Shapley untuk fitur ke- j . Mengingat model $g(\mathbf{z})$ dan nilai Shapley ϕ_j , Penentuan nilai setiap karakteristik input adalah sebagai berikut:

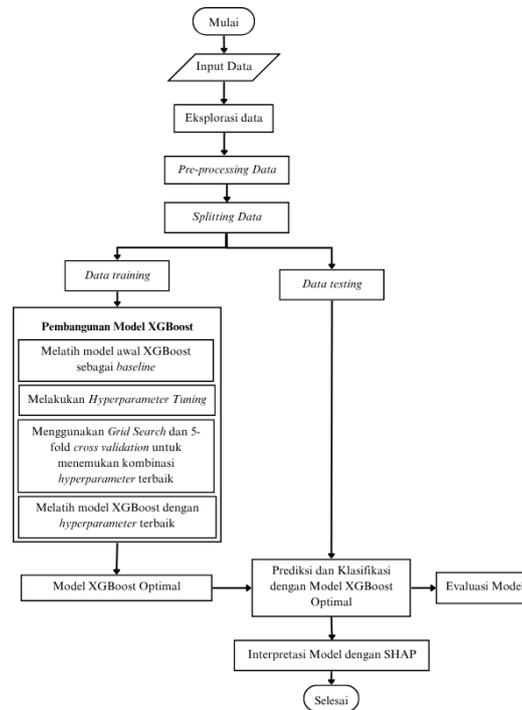
$$\phi_j = \sum_{S \subseteq N \setminus \{j\}} \frac{|S|!(M-|S|-1)!}{M!} [f(S \cup \{i\}) - f(S)] \quad (4)$$

Di mana $\phi_j \in R$ adalah nilai shapley untuk fitur ke- j , S adalah subset dari jumlah prediktor, $f(S)$ adalah nilai fungsi karakteristik setiap prediktor dan N adalah jumlah total fitur yang tersedia.



Gambar 1. Ilustrasi SHAP

Selain perhitungan matematis, SHAP juga menyediakan visualisasi untuk memahami peran masing-masing fitur (Gambar 1). Misalnya, *summary plot* dapat digunakan untuk melihat distribusi dan ukuran *Shapley* secara keseluruhan, *dependence plot* berguna untuk memahami hubungan *Shapley* dan nilai fitur, sedangkan *force plot* dapat menunjukkan bagaimana kombinasi fitur pada sebuah observasi turut mendorong prediksi naik atau turun. Dengan pendekatan dan visualisasi tersebut, interpretasi model menjadi lebih rinci, transparan, dan mudah dipahami, sehingga berguna baik bagi ahli data, pengambil keputusan, maupun kalangan non-teknis (Hadi et al., 2023).



Gambar 2. Alur Penelitian

Tahapan analisis pada diagram alir tersebut:

1. Mulai – Input Data
Memulai proses dengan menginput data yang akan dianalisis, yaitu data pasien diabetes dari *Dataset Early Stage Risk Prediction*.
2. Eksplorasi Data
Mengevaluasi distribusi, nilai yang hilang, outlier, dan hubungan antar variabel secara deskriptif.
3. *Pre-processing Data*
Melakukan pembersihan data, transformasi fitur, encoding variabel kategorik, dan normalisasi jika diperlukan.
4. *Splitting Data*
Membagi data menjadi data *training* dan data *testing* (70%:30%) untuk pelatihan dan evaluasi model.
5. Pembangunan Model XGBoost
 - Melatih model XGBoost sederhana tanpa tuning untuk melihat performa dasar.
 - *Hyperparameter Tuning*: Menyesuaikan parameter model untuk meningkatkan performa.
 - Mencari kombinasi parameter terbaik dengan *Grid Search* dan validasi silang (*5-fold cross-validation*).
 - Model Terbaik: Melatih ulang model dengan parameter optimal.
6. Prediksi dan Klasifikasi
Menggunakan model XGBoost optimal untuk memprediksi hasil pada data testing.
7. Evaluasi Model



Mengukur performa model menggunakan metrik seperti *accuracy*, *precision*, *recall*, dan *F1-score*.

8. Interpretasi Model dengan SHAP

Menjelaskan pengaruh fitur terhadap hasil prediksi secara individual maupun global.

2.4. Data Preprocessing

Persiapan data dilakukan dalam beberapa tahap, yaitu:

1. Pemeriksaan terhadap seluruh variabel, tidak ditemukan adanya nilai kosong, sehingga tidak diperlukan tindakan imputasi atau penghapusan data.
2. Melakukan proses *label encoding* untuk mengubah data kategorik menjadi numerik untuk seluruh variabel prediktor selain variabel usia.
3. Melakukan standarisasi untuk variabel usia yang bertipe numerik untuk menyamakan skala antara fitur numerik dan kategorik.

2.5 Splitting Data

Data yang ada kemudian dibagi menjadi dua bagian, yaitu data latih (*training*) sebesar 70% dan data uji (*testing*) sebesar 30%. Pada penelitian ini terdapat 520 data pasien diabetes, sehingga jumlah data latih dan data uji berturut-turut adalah 364 dan 156 yang kemudian akan dimodelkan dan diukur performanya untuk melihat seberapa baik model telah dibentuk.

2.6 Hypeparameter Tuning

Penyesuaian hyperparameter merupakan salah satu tahapan penting dalam pengembangan model *machine learning* karena dapat mempengaruhi akurasi dan performa model. *Hyperparameter* adalah parameter yang nilainya ditentukan sebelum proses pelatihan dan tidak dipelajari secara langsung dari data. Beberapa hypeparameter dalam XGBoost adalah sebagai berikut:

Tabel 1. Hyperparameter pada metode XGBoost

Parameter	Kegunaan
<i>n_estimators</i>	Jumlah pohon yang digunakan dalam proses klasifikasi
<i>max_depth</i>	Kedalaman maksimum dari pohon keputusan
<i>learning_rate</i>	Penyusutan ukuran langkah dalam pembaruan untuk menghindari overfitting
<i>subsample</i>	Proporsi data yang digunakan saat proses pelatihan
<i>colsample_bytree</i>	Rasio subsample kolom saat membangun setiap pohon

Untuk memperoleh konfigurasi optimal, digunakan metode *Grid Search* yaitu pendekatan yang mengevaluasi berbagai kombinasi nilai *hyperparameter* secara sistematis untuk menemukan konfigurasi terbaik berdasarkan metrik performa tertentu. Meskipun metode ini memerlukan waktu komputasi yang lebih besar, *Grid Search* menawarkan keunggulan dalam cakupan eksplorasi parameter dan kemudahan implementasi. Dengan pendekatan ini, diperoleh kombinasi parameter terbaik yang memaksimalkan kinerja klasifikasi model XGBoost.

2.7 Evaluasi Model

Untuk mengukur seberapa baik kualitas klasifikasi model, dilakukan penyusunan *confusion matrix* berdasarkan perbandingan antara label aktual dan hasil prediksi. Matriks ini memiliki empat komponen penting, yaitu [15]:

- *True Positive* (TP): jumlah kasus yang berhasil diprediksi benar
- *True Negative* (TN): jumlah kasus negatif yang berhasil diprediksi dengan benar

- *False Positive* (FP): jumlah kasus negatif yang diprediksi sebagai positif
- *False Negative* (FN): jumlah kasus positif yang diprediksi sebagai negatif

Nilai-nilai tersebut menjadi dasar dalam menghitung berbagai metrik evaluasi performa model, yang dirumuskan sebagai berikut:

- *Accuracy*

$$Accuracy = \frac{(TP+TN)}{(TP+FP+FN+TN)} \times 100\% \quad (5)$$

- *Precision*

$$Precision = \frac{TP}{(TP+FP)} \times 100\% \quad (6)$$

- *Recall*

$$Recall = \frac{TP}{(TP+FN)} \times 100\% \quad (7)$$

- *F1-Score*

$$F1 - Score = 2 \times \frac{precision \times recall}{precision + recall} \quad (8)$$

III. HASIL DAN PEMBAHASAN

3.1 Preprocessing Data

Tahap awal dalam analisis dilakukan dengan *preprocessing* data untuk memastikan bahwa dataset dalam kondisi bersih dan siap diproses oleh algoritma. Berdasarkan hasil pemeriksaan, dataset tidak mengandung missing value, sehingga tidak diperlukan proses imputasi data.

Dataset terdiri atas 15 fitur kategorik dan 1 fitur numerik (*Age*). Untuk memungkinkan pemrosesan oleh model, seluruh variabel kategorik dikonversi ke bentuk numerik menggunakan teknik label encoding, di mana setiap kategori diubah menjadi nilai integer. Proses encoding ini penting agar XGBoost dapat membangun pohon keputusan berdasarkan nilai-nilai numerik yang merepresentasikan kategori.

3.2 Statistik Deskriptif

Setelah melalui *preprocessing*, dilakukan analisis deskriptif untuk memahami karakteristik dasar data. Dataset ini terdiri atas 520 observasi yang memuat 16 variabel, yang terdiri atas 15 variabel kategorik dan 1 variabel numerik (*Age*). Variabel target berupa status risiko diabetes (positif atau negatif).

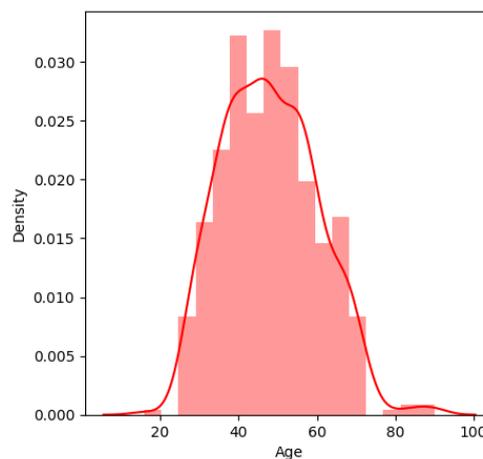
Tabel 2. Distribusi Frekuensi Variabel Kategorik

Variabel	Kategori	Frekuensi	Persentase
<i>Gender</i>	Lak-Laki	328	63,08%
	Perempuan	192	36,92%
<i>Polyuria</i>	Ya	258	49,62%
	Tidak	262	50,38%
<i>Polydipsia</i>	Ya	233	44,81%
	Tidak	287	55,19%
<i>Sudden Weight Loss</i>	Ya	217	41,73%
	Tidak	303	58,27%
<i>Weakness</i>	Ya	305	58,65%
	Tidak	215	41,35%
<i>Polyphagia</i>	Ya	237	45,58%
	Tidak	283	54,42%
<i>Genital Thrush</i>	Ya	116	22,31%



	Tidak	404	77,69%
<i>Visual Blurring</i>	Ya	233	44,81%
	Tidak	287	55,19%
<i>Itching</i>	Ya	253	48,65%
	Tidak	267	51,35%
<i>Irritability</i>	Ya	126	24,23%
	Tidak	394	75,77%
<i>Delayed Healing</i>	Ya	239	45,96%
	Tidak	281	54,04%
<i>Partial Paresis</i>	Ya	224	43,08%
	Tidak	296	56,92%
<i>Muscle Stiffness</i>	Ya	195	37,50%
	Tidak	325	62,50%
<i>Alopecia</i>	Ya	179	34,42%
	Tidak	341	65,58%
<i>Obesity</i>	Ya	88	16,92%
	Tidak	432	83,08%
<i>Class</i>	Positif	320	61,54%
	Negatif	200	38,46%

Berdasarkan Tabel 2, terlihat bahwa proporsi pasien dengan risiko diabetes (positif) mencapai 61,54%, lebih tinggi dibandingkan dengan yang negatif. Gejala umum seperti *weakness*, *polyuria*, *itching*, *delayed healing*, dan *polyphagia* muncul pada lebih dari 45% responden, sedangkan gejala seperti *obesity* dan *genital thrush* memiliki frekuensi relatif lebih rendah. Komposisi jenis kelamin menunjukkan dominasi responden laki-laki dengan persentase 63,08%.



Gambar 3. Distribusi Variabel “Age”

Berdasarkan Gambar 3, Distribusi usia membentuk pola menyerupai distribusi normal dengan konsentrasi terbesar pada rentang usia 40–60 tahun. Nilai ekstrem sangat jarang ditemukan, dengan sebagian besar responden berusia antara 25 hingga 70 tahun. Distribusi ini menunjukkan bahwa risiko diabetes lebih banyak ditemukan pada usia dewasa hingga lanjut usia.

3.3 Hyperparameter Tuning

Proses tuning parameter dilakukan untuk memperoleh kombinasi parameter XGBoost yang menghasilkan performa terbaik. Teknik *grid search* digunakan dengan *5-fold cross-validation* untuk menghindari *overfitting* terhadap data latih.



Tabel 3. Nilai Parameter yang Dicoba dan Kombinasi Terbaik XGBoost

Parameter	Nilai	Nilai Terbaik
<i>learning_rate</i>	0.01, 0.1, 0.3	0.3
<i>n_estimators</i>	50, 100, 200	200
<i>max_depth</i>	3, 5, 7	3
<i>subsample</i>	0.7, 0.8, 0.9	0.9
<i>colsample_bytree</i>	0.7, 0.8, 0.9	0.7

3.4 Klasifikasi XGBoost

Setelah model dilatih menggunakan parameter optimal, model XGBoost dibangun menggunakan 70% data latih. Model kemudian diuji menggunakan data uji (30% dari total data) untuk mengevaluasi performanya. Hasil prediksi ditampilkan dalam confusion matrix pada Tabel 4.

Tabel 4. Confusion Matrix

Aktual	Prediksi	
	Negatif	Positif
Negatif	54	0
Positif	2	100

Tabel 4 menyajikan hasil klasifikasi pada data uji menggunakan algoritma XGBoost. Berdasarkan hasil tersebut, dapat dilihat bahwa dari seluruh data aktual berkategori positif diabetes, sebanyak 100 data berhasil diprediksi dengan benar sebagai positif. Namun, terdapat 2 data positif yang salah diklasifikasikan oleh model sebagai negatif, sehingga termasuk dalam kategori *false negative*. Sementara itu, dari data yang sebenarnya berkategori negatif, sebanyak 54 data berhasil diprediksi dengan benar sebagai negatif. Tidak terdapat data positif yang salah diklasifikasikan sebagai negatif, sehingga nilai *false positive* adalah nol. Hal ini menunjukkan bahwa model memiliki kemampuan yang sangat baik dalam mengenali kasus negatif, serta hanya mengalami kesalahan minor dalam mengklasifikasikan sebagian kecil data positif.

Tabel 5. Hasil Metriks Evaluasi Model XGBoost

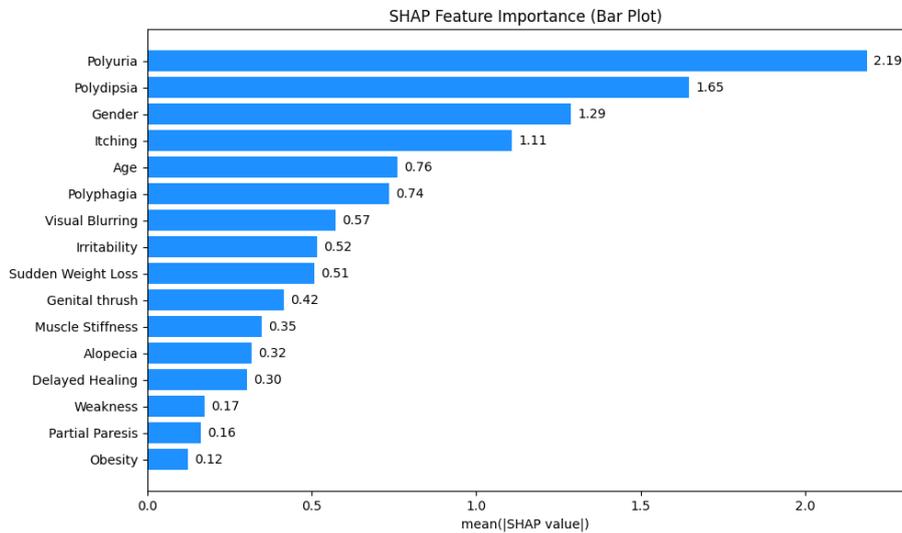
Metriks Evaluasi	<i>Data Train</i>	<i>Data Test</i>
<i>Accuracy</i>	100%	98,7%
<i>Precision</i>	100%	100%
<i>Recall</i>	100%	98%
<i>F1-Score</i>	100%	99%

Berdasarkan Tabel 5, evaluasi lebih lanjut dilakukan menggunakan empat metrik utama. Akurasi model, yang mengukur proporsi prediksi benar terhadap keseluruhan data, mencapai 98,70%. Presisi, yang menunjukkan ketepatan prediksi pada kelas positif, bernilai 100%, sedangkan recall, yaitu kemampuan model dalam menangkap seluruh kasus positif, mencapai 98% persen. Nilai F1-score sebagai harmonisasi antara presisi dan recall juga sangat tinggi, yaitu 99%. Secara keseluruhan, hasil evaluasi ini menunjukkan bahwa model XGBoost memiliki performa yang sangat baik dalam klasifikasi risiko diabetes, terutama dalam mendeteksi seluruh kasus positif tanpa kesalahan.

3.5 Interpretasi Model dengan SHAP

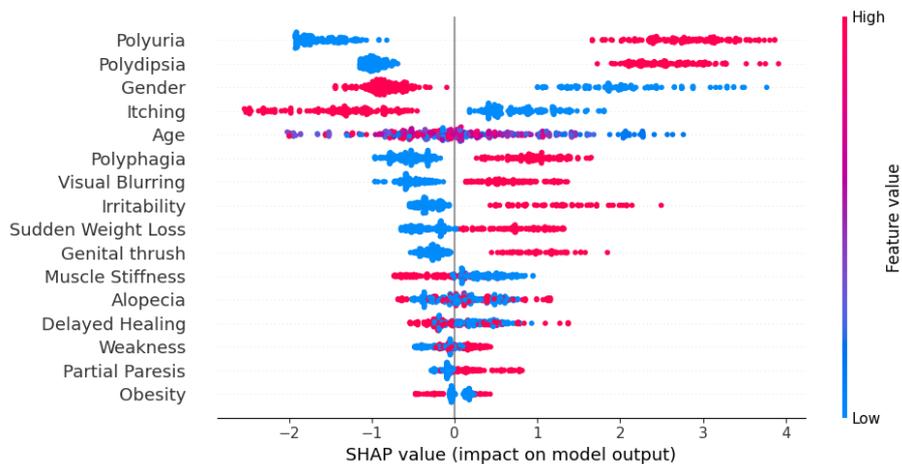
Untuk memahami bagaimana setiap fitur berkontribusi terhadap prediksi model, dilakukan analisis interpretatif menggunakan SHAP (*Shapley Additive Explanations*). Dua visualisasi utama

digunakan dalam interpretasi ini, yaitu SHAP bar plot dan SHAP summary plot, seperti ditampilkan pada Gambar 4 dan Gambar 5.



Gambar 4. Visualisasi SHAP Bar Plot

SHAP bar plot menunjukkan urutan fitur berdasarkan rata-rata nilai absolut SHAP. Semakin besar nilai SHAP rata-rata suatu fitur, semakin besar pengaruh fitur tersebut terhadap Keputusan klasifikasi secara keseluruhan. Berdasarkan Gambar 4, terlihat bahwa fitur *polyuria* menempati urutan pertama dengan nilai rata-rata SHAP sebesar 2,19, yang menunjukkan bahwa fitur ini secara konsisten memiliki dampak paling besar dalam mendorong model mengklasifikasikan responden sebagai penderita diabetes. Selanjutnya, *polydipsia* memiliki nilai 1,65, dan *gender* sebesar 1,29, yang juga menandakan kontribusi penting terhadap hasil prediksi. Fitur lain seperti *itching* (1,11) dan *age* (0,76) memiliki pengaruh yang moderat, sedangkan fitur-fitur seperti *obesity* (0,12), *partial paresis* (0,16), dan *weakness* (0,17) menunjukkan kontribusi yang rendah terhadap prediksi model.



Gambar 5. Visualisasi SHAP Summary Plot

Selanjutnya, SHAP summary plot memberikan informasi mengenai sebaran nilai SHAP untuk setiap fitur, yang menggambarkan arah dan intensitas kontribusinya terhadap hasil prediksi.



Visualisasi ini menggambarkan sebaran nilai SHAP untuk setiap fitur pada seluruh observasi. Setiap titik mewakili satu data, dengan warna yang menunjukkan nilai fitur tersebut: titik berwarna merah menunjukkan nilai fitur yang tinggi, sedangkan titik berwarna biru menunjukkan nilai yang rendah. Titik-titik yang berada di sisi kanan sumbu nol menunjukkan bahwa fitur tersebut memberikan kontribusi ke arah prediksi positif diabetes, sementara titik di sisi kiri berkontribusi terhadap prediksi negatif.

Berdasarkan Gambar 5, Pada fitur *polyuria* dan *polydipsia*, terlihat bahwa nilai yang tinggi (warna merah) secara konsisten terletak di sisi kanan, yang menunjukkan bahwa responden yang mengalami kedua gejala tersebut cenderung diklasifikasikan sebagai penderita diabetes. Fitur *gender* menunjukkan bahwa nilai tertentu memiliki kecenderungan memengaruhi prediksi model. Titik-titik biru pada fitur ini, yang merepresentasikan kategori "perempuan", cenderung berada di sisi kanan, mengindikasikan bahwa model mengasosiasikan jenis kelamin perempuan dengan peningkatan risiko diabetes dalam data ini. Sebaliknya, titik merah (laki-laki) cenderung tersebar lebih ke arah sisi kiri atau netral. Hal ini menunjukkan bahwa dalam model, perempuan memiliki kecenderungan lebih tinggi untuk diklasifikasikan sebagai positif diabetes dibandingkan laki-laki. Sementara itu, fitur *age* juga menunjukkan pola yang jelas. Titik merah yang menandakan usia tinggi berada di sisi kanan, mengindikasikan bahwa semakin tinggi usia seseorang, semakin besar kontribusinya terhadap prediksi sebagai penderita diabetes. Sebaliknya, nilai usia yang rendah (titik biru) umumnya berada di sisi kiri, mendorong prediksi ke arah kelas negatif. Pola ini sesuai dengan pengetahuan medis bahwa risiko diabetes meningkat seiring bertambahnya usia.

Dari keseluruhan visualisasi, dapat disimpulkan bahwa gejala klasik seperti *polyuria* dan *polydipsia* tetap menjadi indikator utama risiko diabetes, namun variabel demografis seperti *gender* dan *age* juga memberikan pengaruh penting dalam pengambilan keputusan model.

IV. KESIMPULAN

Penelitian ini bertujuan untuk membangun model klasifikasi risiko diabetes dan mengidentifikasi faktor-faktor yang paling berkontribusi terhadap hasil prediksi dengan menggunakan algoritma XGBoost dan metode interpretasi SHAP. Hasil menunjukkan bahwa model XGBoost mampu mengklasifikasikan data dengan sangat baik, ditunjukkan oleh nilai *accuracy* sebesar 98,7%, *precision* 100%, *recall* 98%, serta *F1-score* 99%. Analisis interpretatif menggunakan SHAP menunjukkan bahwa fitur *polyuria* dan *polydipsia* merupakan variabel yang paling berpengaruh dalam menentukan hasil klasifikasi. Selain itu, variabel *gender* dan *age* juga memberikan kontribusi penting, di mana jenis kelamin laki-laki dan usia yang lebih tinggi cenderung meningkatkan kemungkinan prediksi positif diabetes. Secara keseluruhan, kombinasi XGBoost dan SHAP tidak hanya mampu menghasilkan prediksi yang akurat, tetapi juga memberikan interpretasi yang jelas terhadap pengambilan keputusan model. Pendekatan ini dapat digunakan sebagai alat bantu yang potensial dalam sistem deteksi dini risiko diabetes, sekaligus mendukung transparansi dalam pengambilan keputusan berbasis data di bidang kesehatan.

REFERENSI

1. I. D. Atlas, *IDF Diabetes Atlas 10th Edition*. 2021. doi: 10.1016/j.diabres.2013.10.013.
2. A. Rosyada and I. Trihandini, "Determinan Komplikasi Kronik Diabetes Melitus pada Lanjut Usia," *Kesmas: National Public Health Journal*, vol. 7, no. 9, p. 395, 2013, doi: 10.21109/kesmas.v7i9.11.



3. E. Safitri, D. Rofianto, N. Purwati, H. Kurniawan, and S. Karnila, “Prediksi Penyakit Diabetes Melitus Menggunakan Algoritma Machine Learning,” *Jurnal Sistem dan Teknologi Informasi*, vol. 12, no. 4, pp. 760–766, 2024, doi: 10.26418/justin.v12i4.84620.
4. M. K. Nasution, Rd. R. Saedudin, and V. P. Widartha, “Perbandingan Akurasi Algoritma Naïve Bayes Dan Algoritma Xgboost Pada Klasifikasi Penyakit Diabetes,” *e-Proceeding of Engineering*, vol. 8, no. 5, pp. 9765–9772, 2021.
5. G. Abdurrahman and M. Sintawati, “Implementation of XGBoost for Classification of Parkinson’s Disease,” *Journal of Physics: Conference Series*, vol. 1538, no. 1, 2020, doi: 10.1088/1742-6596/1538/1/012024.
6. C. E. Sukmawati, A. Fitri, N. Masruriyah, and A. R. Juwita, “Efektivitas algoritma AdaBoost dan XGBoost pada dataset obesitas populasi dewasa,” vol. 6, no. 2, pp. 101–111, 2024, doi: 10.37905/jji.
7. M. Fikri, “Klasifikasi Status Stunting Pada Anak Bawah Lima Tahun Menggunakan Extreme Gradient Boosting,” *Merkurius: Jurnal Riset Sistem Informasi dan Teknik Informatika*, vol. 2, no. 4, pp. 173–184, 2024.
8. M. J. Raihan, M. A. M. Khan, S. H. Kee, and A. Al Nahid, “Detection of the chronic kidney disease using XGBoost classifier and explaining the influence of the attributes on the model using SHAP,” *Scientific Reports*, vol. 13, no. 1, pp. 1–15, 2023, doi: 10.1038/s41598-023-33525-0.
9. T. Kurniawan, L. Hermawanti, and A. N. Safriandono, “Interpretable Machine Learning with SHAP and XGBoost for Lung Cancer Prediction Insights,” vol. 8, no. 2, pp. 296–303, 2024.
10. S. E. Herni Yulianti, Oni Soesanto, and Yuana Sukmawaty, “Penerapan Metode Extreme Gradient Boosting (XGBOOST) pada Klasifikasi Nasabah Kartu Kredit,” *Journal of Mathematics: Theory and Applications*, vol. 4, no. 1, pp. 21–26, 2022, doi: 10.31605/jomta.v4i1.1792.
11. Y. Rombe, S. A. Thamrin, and A. Lawi, “Application of Adaptive Synthetic Nominal and Extreme Gradient Boosting Methods in Determining Factors Affecting Obesity: A Case Study of Indonesian Basic Health Research Survey 2013,” *Indonesian Journal of Statistics and Its Applications*, vol. 6, no. 2, pp. 309–317, 2022, doi: 10.29244/ijsa.v6i2p309-317.
12. R. Sibindi, R. W. Mwangi, and A. G. Waititu, “A boosting ensemble learning based hybrid light gradient boosting machine and extreme gradient boosting model for predicting house prices,” *Engineering Reports*, vol. 5, no. 4, pp. 1–19, 2022, doi: 10.1002/eng2.12599.
13. T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, vol. 13-17-Aug, pp. 785–794, 2016, doi: 10.1145/2939672.2939785.
14. S. M. Lundberg and S. I. Lee, “A unified approach to interpreting model predictions,” *Adv Neural Inf Process Syst*, vol. 2017-Decem, no. Section 2, pp. 4766–4775, 2017.
15. M. Dava Maulana, A. Id Hadiana, and F. Rakhmat Umbara, “Algoritma Xgboost Untuk Klasifikasi Kualitas Air Minum,” *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 7, no. 5, pp. 3251–3256, 2024, doi: 10.36040/jati.v7i5.7308.